Molecular characterisation of the vacuolating autotransporter toxin in uropathogenic escherichia coli

Nichols, Katie B; Totsika, Makrina; Moriel, Danilo G; Lo, Alvin W; Yang, Ji; Wurpel, Daniël J; Rossiter, Amanda E; Strugnell, Richard A; Henderson, Ian R; Ulett, Glen C; Beatson, Scott A; Schembri, Mark A

DOI:
10.1128/JB.00791-15

Molecular characterisation of the Vacuolating Autotransporter Toxin in Uropathogenic *Escherichia coli*

Katie B. Nichols[1], Makrina Totsika[2], Danilo G. Moriel[1], Alvin W. Lo[1], Ji Yang[3], Daniël J. Wurpel[1], Amanda E. Rossiter[3], Richard A. Strugnell[3], Ian R. Henderson[4], Glen C. Ulett[5], Scott A. Beatson[1] and Mark A. Schembri[1*]

[1]*Australian Infectious Disease Research Centre, School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, Queensland, Australia.*

[2]*Institute of Health and Biomedical Innovation, School of Biomedical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia*

[3]*Department of Microbiology and Immunology, The University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia*

[4]*Institute of Microbiology and Infection, School of Immunity and Infection, University of Birmingham, Birmingham, B15 2TT, United Kingdom*

[5]*School of Medical Science, Menzies Health Institute Queensland, Griffith University, Gold Coast, Queensland, Australia*

Running title: Characterisation of Vat in UPEC

Key words: Uropathogenic Escherichia coli, urinary tract infection, autotransporter, serine protease, vacuolating autotransporter toxin.

*** Corresponding author.**

Mailing address: School of Chemistry and Molecular Biosciences, Building 76, University of Queensland, Brisbane QLD 4072, Australia. Phone: +617 33653306; Fax: +617 33654699; E-mail: m.schembri@uq.edu.au

ABSTRACT

The vacuolating autotransporter (AT) toxin (Vat) contributes to Uropathogenic *Escherichia coli* (UPEC) fitness during systemic infection. Here we characterised Vat and investigated its regulation in UPEC. We assessed the prevalence of *vat* in a collection of 45 UPEC urosepsis strains and showed that it was present in 31 (68%) of the isolates. The isolates containing the *vat* gene corresponded to three major *E. coli* sequence types (ST12, 73 and 95) and these strains secreted the Vat protein. Further analysis of the *vat* genomic locus identified a conserved gene located directly downstream of *vat* that encodes a putative MarR-like transcriptional regulator, which we termed *vatX*. The *vat-vatX* genes were present in the UPEC reference strain CFT073 and RT-PCR revealed both genes are co-transcribed. Over-expression of *vatX* in CFT073 led to a 3-fold increase in *vat* gene transcription. The *vat* promoter region contained three putative nucleation sites for the global transcriptional regulator H-NS; thus the *hns* gene was mutated in CFT073 (to generate CFT073*hns*). Western blot analysis using a Vat-specific antibody revealed a significant increase in Vat expression in CFT073*hns* compared to wild-type CFT073. Direct H-NS binding to the *vat* promoter region was demonstrated using purified H-NS in combination with electrophoresis mobility shift assays. Finally, Vat-specific antibodies were detected in plasma samples from urosepsis patients infected by *vat*-containing UPEC strains, demonstrating Vat is expressed during infection. Overall, this study has demonstrated that Vat is a highly prevalent and tightly regulated immunogenic SPATE secreted by UPEC during infection.

IMPORTANCE

Uropathogenic *Escherichia coli* (UPEC) are the major cause of hospital and community acquired urinary tract infections. The Vacuolating autotransporter toxin (Vat) is a cytotoxin known to contribute to UPEC fitness during murine sepsis infection. In this study, Vat was found to be highly conserved and prevalent among a collection of urosepsis clinical isolates, and expressed at human core body temperature. Regulation of *vat* was demonstrated to be directly repressed by the global transcriptional regulator H-NS and upregulated by the downstream gene *vatX* (a new MarR-type transcriptional regulator). Additionally, increased Vat-specific IgG titres were

detected in plasma from corresponding urosepsis patients infected with *vat*-positive isolates. Hence, Vat is a highly conserved and tightly regulated urosepsis-associated virulence factor.

INTRODUCTION

Urinary tract infections (UTIs) are one of the most common human infections, and affect 40-50% of women and approximately 12% of men globally (1). UTIs are ascending infections and can involve infection of the bladder (cystitis), kidneys (pyelonephritis) or dissemination into the bloodstream (urosepsis). Uropathogenic *Escherichia coli* (UPEC) are the primary etiological agent of UTI and cause 70-90 % of all such infections (2). UPEC can survive in the urinary tract and cause disease due to a diverse range of virulence factors including fimbriae (3-6), autotransporter (AT) proteins (7-10), surface polysaccharides such as the O-antigen and capsule (11-13), iron acquisition systems (14-16) and toxins (17-21).

AT proteins constitute a large family of proteins from Gram-negative bacteria that are translocated by a dedicated type V secretion system (reviewed in 22, 23-26). AT translocation also requires accessory proteins including the β-barrel assembly module (BAM) and the translocation and assembly module (TAM) (27-30). AT proteins consist of three major domains: (i) a signal peptide that targets the protein to the secretory apparatus for inner membrane translocation; (ii) a passenger domain that comprises the functional domain of the protein; and (iii) a translocator domain that inserts into the outer membrane (reviewed in 22, 23, 25, 31-33) . One major subgroup of AT proteins is the serine protease AT proteins of *Enterobacteriaceae* (SPATEs). SPATEs are characterised by the presence of an immunoglobulin A1-like protease domain (PF02395) within the passenger domain that contains the conserved serine protease motif GDSGS (34, 35). The first serine within this motif comprises the catalytic triad in conjunction with upstream conserved histidine and aspartate residues. SPATEs can be phylogenetically grouped into two classes (reviewed in 34, 36, 37). Class I SPATEs represent the major group of these proteins and exhibit cytotoxic activity (37-43). Class II SPATEs recognise a more diverse range of substrates including mucins (reviewed in 34, 36, 37) and immunomodulatory host proteins (44).

The vacuolating AT toxin (Vat) of *E. coli* is a class II SPATE (34, 36, 45) that exhibits cytotoxicity to chicken embryonic fibroblast cells and contributes to avian cellulitis infection (46). The *vat* gene was originally identified within a pathogenicity island (Vat-PAI) from the

avian pathogenic *E. coli* (APEC) strain Ec222 (46). The Vat-PAI is integrated into the Ec222 chromosome at the *thrW*-tRNA site between the *proA* and *yagU* genes (45, 46). The Vat-PAI from Ec222 consists of 33 open reading frames (ORFs), with the *vat* gene residing at ORF#27. Only five additional ORFs in this PAI were reported to share homology with other previously known protein sequences. This includes the ORF located downstream of *vat* (ORF#26), which shares 44% amino acid identity to the P pilus associated transcriptional regulatory protein PapX from UPEC strain CFT073 (46). PapX belongs to the family of multiple antibiotic resistance (MarR) regulators of *Enterobacteriaceae* and contributes to flagella regulation by binding to the promoter region of the *flhDC* master regulator genes (47-49). In UPEC, the *vat* gene is associated with virulence and contributes to survival during murine systemic infection (50).

The full-length Vat protein is ~140 kDa and is processed during translocation to release a 111.8 kDa passenger domain into the extracellular milieu. Vat shares 78% identity to the APEC associated Temperature-sensitive hemagglutinin (Tsh), which is almost identical (99% amino acid identity) to the SPATE Haemoglobin binding protein (Hbp) (51, 52). Hbp has been analysed extensively in the *E. coli* intra-abdominal clinical strain EB1, and its crystal structure has been solved (53, 54). Tsh/Hbp possess dual proteolytic and adhesive properties (55-57). Unlike Tsh/Hbp, Vat is unable to digest casein at 37°C (45, 46).

Despite these functional differences, the high protein sequence identity shared between Tsh/Hbp and Vat has led to confusion in the annotation of *vat* genes within *E. coli* genomes available on the NCBI database. For example, the CFT073 *vat* gene (c0393) has been annotated as *hbp* (58), and even referred to as *tsh* due to its temperature-dependent regulation (59). In addition, the *vat* gene from UPEC strain 536 is annotated as *sepA*, which encodes the *Shigella* extracellular protein A (45).

In this study, we have examined the sequence conservation of *vat* genes from available *E. coli* genomes and compared its genomic location, with the aim to correct existing annotation errors and *vat* nomenclature. We also examined the role of the putative MarR regulator identified downstream of the *vat* gene as well as the histone-like nucleoid protein H-NS in regulation of the *vat* gene. Finally, we examined the prevalence, expression and secretion of Vat in a collection of

UPEC urosepsis isolates, and investigated its immunogenicity by examining plasma from urosepsis patients.

MATERIALS AND METHODS

Ethics statement. This study was performed in accordance with the ethical standards of The University of Queensland, Princess Alexandra Hospital, Gold Coast Hospital, Queensland Health, Griffith University and the Helsinki Declaration. The study was approved, and the need for informed consent was waived by the institutional review boards of the Princess Alexandra Hospital (2008/264), Queensland Health and Griffith University (MSC/18/10/HREC).

Bacterial strains and growth conditions. *E. coli* strains CFT073 (60), IHE3034 (61), 536 (62), MG1655 (63) and BL21 (64), as well as the *E. coli* reference (ECOR) collection (65), have been described previously. The 45 urosepsis UPEC strains were isolated from the blood of patients presenting with urosepsis at the Princess Alexandra Hospital (Brisbane, Australia). A matching urine sample was also cultured from each patient; in all cases the blood and urine isolates were identical as determined by virulence gene profiling. Unless otherwise stated, strains used in this study were routinely grown at 37°C on solid or in liquid Lysogeny broth (LB) supplemented with antibiotics: kanamycin (kan [100μg/mL]), ampicillin (amp [100μg/mL]) or chloramphenicol (cam [30μg/mL]). Supplementation of growth media with L-arabinose (0.2% [w/v]) or isopropyl β-D-1-thiogalactopyranoside (IPTG [1mM]) was used to induce plasmid-mediated gene expression.

Bioinformatic analysis. The presence of the *vat* gene was determined in 77 complete *E. coli* genomes (listed in Table S1) available from the National Centre for Biotechnology Information (NCBI) database by BLAST analysis using the *vat* gene (c0393) from the CFT073 genome (Genbank accession no.: AE014075.1 (58)) as a search tool. The cut-off was set at >85% amino acid identity of the encoded protein sequence. The genomic location surrounding the *vat* gene in each of the *vat*-positive strains was investigated in Artemis (66). All *vat* genes identified were located on a PAI defined by the *proA* and *yagU* genes. The nucleotide sequence of each *vat*-associated PAI was compared in EasyFig (67).

A comparative protein analysis of the MarR family of transcriptional regulators (Table S2) was performed to analyse their relative phylogenetic relationship to VatX. The MarR dataset was compiled using an iterative approach that involved BLAST analysis against the 77 complete NCBI *E. coli* genomes listed in Table S1. Representative protein sequences, underlined in Table S2, were chosen for each MarR type regulator based on previous characterisation in the literature. These sequences included MarR from MG1655 (b1530), MprA(EmrR) from MG1655 (b2684), HosA from E2348/69 (E2348C_3010), HpcR/ HpaR from strain W (WFL_22965), SlyA from MG1655 (b1642) and PapX from CFT073 (c3582). Each of the representative sequences were used to BLAST against the 77 complete *E. coli* genomes and 330 homologous protein sequences were identified (E<0.001). The evolutionary relationship between VatX and other representative MarR regulators, as well as the protein sequences listed in Table S2, was inferred using ClustalΩ (68, 69) and visualised through FigTree (70).

DNA manipulation and genetic techniques. DNA techniques were performed as previously described (71). Isolation of plasmid DNA was performed using the QIAprep spin column miniprep kit (QIAGEN). Polymerase chain reactions (PCR) were performed using the specified primers which were sourced from Integrated DNA Technologies (Singapore). PCR products were amplified using *Taq* DNA polymerase according to the manufacturer's instructions (New England Biolabs). Sequencing reactions were performed using the BigDye Terminator v3.1 cycle DNA sequencing kit as per the manufacturer's specifications (Applied Biosystems) and analysed by the Australian Equine Genome Research Centre. Cloning reactions involving restriction endonucleases were performed as per the manufacturer's instructions (New England Biolabs).

Multi locus sequence typing (MLST) and PCR screening. Prevalence of the *vat* gene was assessed by PCR using primers 2020 (5'-GTATATGGGGGGCAACATAC-3') and 2021 (5'-GTGTCAGAACGGAATTGTCG-3'), which were designed based on the sequence of the *vat* gene from CFT073 (c0393). The *vat* gene sequence from ten of the 31 *vat*-positive UPEC urosepsis strains was determined and deposited on the NCBI database (accession numbers: PA11B *vat*, KR094926; PA15B *vat*, KR094927; PA32B *vat*, KR094928; PA38B *vat*, KR094929; PA42B *vat*, KR094930; PA48B *vat*, KR094931; PA56B *vat*, KR094932; PA57B *vat*, KR094933; PA60B *vat*, KR094934; PA66B *vat*, KR094935). The sequence type of the

UPEC urosepsis strains was determined using the seven-gene MLST scheme (http://mlst.ucc.ie/mlst/dbs/Ecoli) (72). PCR was performed as follows: initial denaturation at 94°C for 5 m; 25 cycles of denaturation at 94°C for 30 s, annealing at 50°C for 30 s, extension at 72°C for 30 s followed by a final extension at 72°C for 7 m.

Construction of deletion mutants. The *vat* (c0393), *vatX* (c0392) and *hns* (c1701) genes were mutated in CFT073 using λ-Red mediated homologous recombination (73). Briefly, the kanamycin gene from pKD4 or the chloramphenicol gene from pKD3 were amplified using PCR primers containing 50-bp flanking regions homologous to the target genes *vat* (3353: 5'-TCGTAATGAACACAGTTCATCTGATCTCCACACACCAAGACTTGATAAGCTcacgtcttgagcgattgtgtagg-3' and 3354: 5'-GAAACCACCACCCCATGATTTTGTTTTACCGCTGTACAGGCCTGCTGACGCgacatgggaattagccatggtcc-3'), *vatX* (5232: 5'-TTCACGATACTTCATGTAACACTCAGGTTGAGTAATCTTCgtgtaggctggagctgcttc-3' and 5233: 5'-AGAATACATTGTAAGAAGATGACTGTTAGTATGTTTTAACAcatatgaatatcctcctta-3') or *hns* (1583: 5'-TCGTGCGCAGGCAAGAGAATGTACACTTGAAACGCTGGAAGAAATGCTGGgtgtaggctggagctgcttc-3' and 1584: 5'-TTGATTACAGCTGGAGTACGGCCCTGGCCAGTCCAGGTTTTAGTTTCGCCcatatgaatatcctccttag-3'). Amplified fragments were transformed into CFT073(pKD56) expressing of the λ-Red recombinase in order to facilitate homologous recombination for inactivation deletion of the target gene. Removal of the antibiotic resistance gene cassette was performed using plasmid pCP20 as previously described, and enabled the construction of the CFT073*vatX hns* double mutant.

Construction of plasmids. A segment of the *vat* gene corresponding to amino acid residues 63-465 of the passenger domain was amplified from CFT073 using primers 1491 (5'-tacttccaatccaatgcTCCTTACCAGACATACCGCG-3') and 1494 (5'-ttatccacttccaatgTTACCCCGCATATTGATCATTGCC-3') and cloned as a 6 x histidine N-terminal fusion into the pLicA vector using ligation independent cloning (designated pVat^{403}).

The full-length *vat* gene (c0393) and the downstream gene *vatX* gene (c0392) were PCR amplified from CFT073 using the following primer pairs; *vat*: 1524 (5'-cgcgCTCGAGataataaggaattactATGAATAAAATATACGCTC-3') and 1525 (5'-cgcgcaagcttCAAAGCAATAGTCCCTTTGC-3'); and *vatX*: 5244 (5'-cgcgctcgagataataaggaaTCTTCATGAGTTTTCTTTTGCCGTGTGG-3') and 5245 (5'-cccggaagcttTCAATTAACATTAAGGTTTGATA-3'). The PCR products were purified and cloned into XhoI-HindIII digested pSU2718 to generate the plasmids pVat and pVatX. Transcription of the *vat* and *vatX* genes in these plasmids was regulated by the *lac* promoter (74).

Comparative quantitative reverse transcriptase PCR (qRT-PCR). Comparative qRT-PCR was performed essentially as previously described (47). Briefly, strains CFT073, CFT073*vatX* and CFT073*vatX* (pVatX) were grown in LB broth (supplemented with IPTG) until exponential growth phase. The total RNA from each strain was extracted using the RNeasy mini kit as per manufacturer's instructions (QIAGEN). Samples were subjected to RNase free DNA digestion and first strand cDNA synthesis was performed using SuperscriptIII (Invitrogen Life Technologies) with random hexamer (50ng/μL) primers (Invitrogen Life Technologies). Residual RNA was digested by RNaseH and samples were re-purified as recommended by the manufacturer (QIAGEN). The ViiA 7 instrument and software (v 1.2.1) was used to carry out RT PCR reactions (95°C 10 s; 95°C 15 s, 60°C 15 s and 72°C 30 s for 40 cycles). Primers specific to the *vat* gene (5470: 5'-TACCGTAACCAGCTCATCAACAG-3' and 5471: 5'-CATACCCACCTGTTACCCAATGT-3') and *gapA* (control; 820: 5'-GGTGCGAAGAAAGTGGTTATGAC-3' and 821: 5'-GGCCAGCATATTTGTCGAAGTTAG-3') were used to amplify transcripts with SybrGreenI (5 μL) master mix (Applied Biosystems). Each reaction was performed in triplicate and a subsequent melt curve was generated for validation of the results (95°C 15 s, 60°C 1 m and 95°C for 10 s). Cycle threshold values obtained were normalised to the endogenous control and the $2^{-\Delta\Delta Ct}$ method (75) was applied for the comparative analysis. The resulting ratios were statistically analysed using a one-way ANOVA. All experiments were performed in triplicate.

5' RACE and Virtual Footprint analysis. The transcriptional start site for *vat* was determined using the 5' RACE system for rapid amplification of cDNA ends (version 2.0, Invitrogen Life

Technologies) following the manufacture's specifications. Two gene specific primers (5863: 5'-ATGCAGATAGTGCCAGAG-3' and 5864: 5'-CTCTGCGGGTACTCCCTTTAC-3') were used. Putative DNA binding motifs in the *vat* promoter region were identified using Virtual Footprint software (76).

Electrophoretic mobility shift assay (EMSA). EMSA was performed essentially as described previously (77) but with minor adaptations. Briefly, four individual fragments (152 bp, 218 bp, 312 bp and 479 bp) were PCR amplified from the plasmid pBR322 (152 bp: 5'-CATTGGACCGCTGATCGT-3' and 5'-CTTCCATTCAGGTCGAGGT-3'; 218 bp: 5'-AATATTATTGAAGCATTTATCAGGGTTA-3' and 5'-ATGATAAGCTGTCAAACATGAGA-3'; 312 bp: 5'-TATCGACTACGCGATCATGG-3' and 5'-TCTCCCTTATGCGACTCCTG-3'; and 479 bp: 5'-GACCGATGCCCTTGAGAG-3' and 5'-GATCGAAGTTAGGCTGGTAAGA-3'). The 218-bp fragment containing the H-NS repressed *bla* gene promoter was included in the assay as a positive control, while the remaining three fragments do not bind H-NS. The *vat* gene promoter region (252 bp) encompassing all three of the putative H-NS binding sites identified, was also PCR amplified (6103: 5'-CCTGAGAAAAAGCAAACAACA-3' and 6104: 5'-TTTTAGAGCGTATATTTTATTCAT-3') from the genomic DNA of CFT073. This 252-bp fragment was added in an equimolar ratio with the control fragments (7.5 nM per fragment [~100 ng]). Purified native H-NS protein was added to each reaction in increasing concentrations (0 μM, 0.1 μM, 0.5 μM and 1.0 μM). Reactions were incubated at room temperature (15 m in H-NS binding buffer to allow for protein-DNA complex formation. Samples were examined by high-resolution agarose gel electrophoresis (3% Lonza Metaphor [50 V at 4°C]), and viewed under ultraviolet light after staining with ethidium bromide (0.5μg/ mL). Invitrogen's 1 kbp+ ladder was used as a molecular marker.

Preparation of supernatant proteins. Bacterial cultures (10 mL) were standardised to an optical density at 600nm equal to 1.0 (OD_{600} = 1.0), centrifuged (2057 x *g*), and the supernatant was collected and filtered (0.22 μm). Proteins were precipitated by the addition of 10% trichloroacetic acid (TCA) overnight at 4°C. Following precipitation, supernatant fractions were concentrated by centrifugation (12,100 x *g*) and washed twice with 80% acetone to remove residual TCA. Proteins were resuspended in a final volume of 0.1 mL (100-fold concentration).

Purification of denatured His-tagged Vat protein. A bacterial culture (200mL) of *E. coli* BL21 λDE3 expressing the truncated Vat^{403} protein encoded on plasmid $pVat^{403}$ was grown in LB. Bacterial cells were pelleted by centrifugation (2057 x *g*) and lysed (7M urea, 0.1 M NaH_2PO_4, 0.01 M Tris·Cl [pH 8.0]). The recombinant Vat^{403} protein was purified under denaturing conditions using QIAGEN's Ni-NTA spin column kit. The cleared lysate was passed through a pre-equilibrated column via centrifugation (270 x *g*) to allow for the 6xHis tagged-Vat protein to bind. The column was washed (0.1 M NaH_2PO_4, 0.01 M Tris·Cl ([pH 6.3]) and the bound Vat protein was eluted (0.1 M NaH_2PO_4, 0.01 M Tris·Cl [pH 4.5]) by centrifugation (890 x *g*). Protein concentrations were determined using the bicinchoninic acid protein assay kit as per the manufacturer's instructions (Thermo Scientific Pierce Biotechnology). Purity of the eluted protein was validated by sodium-dodecyl-disulfide polyacrylamide gel electrophoresis (SDS-PAGE) analysis (12% polyacrylamide gel) and Coomassie staining.

Immunoblotting. The purified His-tagged recombinant Vat protein was used to generate a Vat-specific polyclonal antibody following a standard protocol (Institute of Medical and Veterinary Science, South Australia). Concentrated supernatant proteins were re-suspended in 50 µL of SDS loading buffer (100 mM Tris-HCl, 4% w/v SDS, 20% w/v glycerol, 0.2% w/v bromophenol blue, pH 6.8) and a 10 µL sample was boiled for 10 min prior to SDS-PAGE. SDS-PAGE and transfer of proteins to a PVDF membrane for western blot analysis was performed as previously described (78). Anti-Vat polyclonal antibodies were used as the primary antibody, and alkaline phosphatase conjugated anti-rabbit antibodies (Sigma Aldrich) were used as the secondary antibody. SIGMAFAST™ BCIP®/NBT (Sigma-Aldrich) was used as the substrate for detection.

Human plasma samples and measurement of Vat immunogenicity. Blood plasma (collected within 4 days post-admission) and matching clinical isolates were obtained from 45 urosepsis patients admitted to the Princess Alexandra Hospital (Brisbane, Australia). The clinical strains isolated from each urosepsis patient were grouped as Vat positive (Vat+) and Vat-negative (Vat-) according to the prevalence of the *vat* gene, as determined by PCR screening using *vat* specific primers. A negative control group of plasma samples was independently obtained from 42 healthy volunteers with no recent history of UTI. The ELISA assay was performed using Nunc

MaxiSorp flat-bottom 96 well microtiter plates (Thermo Scientific). Each well was coated with recombinant Vat protein (10 µg/ml) using carbonate coating buffer (18 mM Na_2CO_3, 450 mM $NaHCO_3$, pH 9.3 [4°C, overnight]). Plates were washed twice with 0.05% v/v Tween20 PBS (PBST) and blocked with 5% w/v skim milk in PBST (150 µl) for 90 min at 37°C. Each well was then washed four times with PBST prior to incubation (90 min at 37°C) with individual plasma samples (1:10 dilution). Unbound antibodies were removed by washing with PBST. Peroxidase-conjugated anti-human IgG (1:30,000 dilution in 5% skim milk) was applied as a secondary antibody for detection (incubated at 37°C for 90 min). Plates were washed four times with PBST and bound anti-human IgG was detected using 3,3',5,5'-tetramethylbenzidine as a substrate. Reactions were stopped with 1 M HCl. The absorbance of each well was measured at 450 nm using the Spectramax plus^{384} plate reader via the SoftMax Pro®v5 program. Data obtained was analysed using GraphPrism5 software and a one-way ANOVA statistical analysis was performed.

RESULTS

The *vat* gene is located on a pathogenicity island at a conserved genomic location. The prevalence of *vat* was assessed in 77 complete *E. coli* genomes available on the NCBI database (Table S1). The *vat* gene was identified in 14 of these strains; these included the previously characterised *vat*-positive UPEC strains CFT073 and 536, as well as twelve additional strains from which *vat* has not previously been characterised (APEC O1, NRG 857C, LF82, IHE3034, S88, 83972, PMV-1, clone D i2, clone D i14, ATCC 25922, Nissle 1917 and UM146). In all 12 strains, the *vat* gene was part of a pathogenicity island (PAI) flanked by the *proA* and *yagU* genes relative to the *E. coli* K12 MG1655 chromosome. This genomic location is consistent with the original identification of *vat* in APEC strain Ec222 (46). Closer examination of the genomic context of *vat* revealed that the upstream region (i.e. the *yagU* end) is highly conserved. In contrast, the region downstream of *vat* (i.e. the *proA* end) exhibits extensive variation, with a range of different DNA segments inserted at various positions of the PAI in strains APEC 01, 83972, UM146, 536 and Ec222 (Fig. 1A).

Vat is secreted by several genome sequenced UPEC strains. The secretion of Vat following growth in LB broth at 37°C was assessed from a selection of the *vat*-positive UPEC strains described above (i.e. CFT073, IHE3034 and 536). As a positive control, the *vat* gene from CFT073 was amplified by PCR, cloned into the low copy number expression vector pSU2718 to generate the plasmid pVat, and transformed into *E. coli* MG1655 to generate the recombinant strain MG1655(pVat). Western blot analysis using a Vat-specific antibody detected a single band of ~110 kDa that corresponded to the predicted size of the secreted passenger domain of Vat in the supernatant of MG1655(pVat), but not the vector control strain MG1655(pSU2718) (Fig. 1B). The *vat* gene was also mutated in CFT073 to generate null mutant strain CFT073*vat*. SDS-PAGE and western blot analysis of the supernatant fraction obtained from CFT073 and CFT073*vat* using our Vat-specific antibody identified the secreted Vat passenger domain from CFT073 but not CFT073*vat* (Fig. 1B). Finally, we also detected bands corresponding to the Vat passenger domain in the supernatant fraction prepared from strains IHE3034 and 536. Taken together, our data demonstrate that Vat is expressed and secreted by the genome-sequenced strains CFT073, IHE3034 and 536.

A *marR*-like gene is located immediately downstream of the *vat* gene. We were interested to study the regulation of Vat, and noted a small open reading frame located directly downstream of the *vat* gene in all *vat*-positive strains (Fig. 1A). This gene, which we have termed *vatX*, corresponds to c0392 in CFT073 (58) and ORF#26 in the Vat PAI from Ec222 (46). The VatX protein sequence is highly conserved (99% amino acid identity in the 14 *vat*-positive completely sequenced strains described above) and shares 44% amino acid identity with the CFT073 P pilus-associated transcriptional regulator PapX. Further analysis of VatX revealed it contains a MarR PFAM domain (PF01047) and a helix-turn-helix motif characteristic of DNA binding proteins. To examine the relationship between VatX and other regulator proteins, we generated a dataset comprising previously characterised *E. coli* MarR type regulators (Table S2) (47, 79-82). A multiple sequence alignment using representative regulator protein sequences (Fig. 2) as well as a more detailed phylogenetic analysis of all MarR-like sequences identified in the 77 complete *E. coli* genomes described above (Fig S1) revealed that VatX forms a distinct clade within the MarR regulator family, and is most closely related to the PapX, SfaX and FocX fimbriae-associated regulators (47, 80, 83, 84).

Expression of the *vat* gene is upregulated by VatX. The proximity, orientation and conserved genetic organization of the *vat* and *vatX* genes led us to examine if VatX contributes to the regulation of the *vat* gene. In order to study this, we generated a CFT073 *vatX* mutant (CFT073*vatX*) and examined the transcription of *vat* in CFT073 and CFT073*vatX* using comparative qRT-PCR. In addition, the *vatX* gene from CFT073 was PCR amplified and cloned into the pSU2718 expression vector (to generate the plasmid pVatX) and used to complement the CFT073*vatX* mutant. No significant difference was observed in the level of *vat* mRNA transcribed in CFT073 and CFT073*vatX* following growth in LB broth at 37°C (Fig. 3A). However, the over-expression of VatX in CFT073*vatX* (pVatX) resulted in an approximately 3-fold higher level of *vat* mRNA transcript compared to WT CFT073 (Fig. 3A). To explore the effect of VatX on Vat expression further, we compared the level of Vat secreted into the supernatant fraction by CFT073, CFT073*vatX* and CFT073*vatX*(pVatX) by western blot analysis (Fig 3B). Consistent with our transcriptional data, the over-expression of VatX in CFT073*vatX*(pVatX) resulted in a significantly increased level of Vat in the culture supernatant, while no difference in the level of secreted Vat was observed in CFT073 and CFT073*vatX*. A similar increase in secreted Vat was also observed when WT CFT073 was transformed with plasmid pVatX (i.e. strain CFT073[pVatX]) (Fig. 3B). Taken together, our results demonstrate that while deletion of *vatX* does not alter the level of Vat secretion, its over-expression significantly enhances Vat expression.

Transcription of the *vat* gene is directly repressed by H-NS. Given the regulatory effect exhibited by VatX on *vat* transcription, we investigated the promoter region of the *vat* gene to identify putative binding sites for other transcription factors. The transcriptional start site for *vat* was determined using 5' RACE and was mapped to a position 80-bp upstream of the Vat ATG start codon. Consensus -35 (5'-ATCACA-3') and -10 (5'-ATTAAT-3') promoter sequence elements, separated by an 18-bp spacer region, were identified upstream of this site (Fig. 4A). Virtual footprint software was used to analyse the *vat* promoter region for putative regulatory binding sites. From this *in silico* analysis, two putative H-NS nucleation sites were identified on the anti-sense strand overlapping the 18-bp spacer region and the 5' end of the -35 element. A

400 third H-NS nucleation site was determined on the direct strand 10-bp downstream of the
401 transcriptional start site.
402
403 The global transcriptional regulator H-NS is known to bind to curved and A-T rich DNA
404 sequences upstream of several defined UPEC virulence genes (85), including genes encoding for
405 toxins (86-89) and autotransporter proteins (8, 10, 90). To investigate the effect of H-NS on *vat*
406 transcription, the level of Vat expression was compared by western blot analysis of supernatant
407 fractions prepared from WT CFT073, CFT073*vat*, CFT073*vatX*, CFT073*hns* and a CFT073*vatX*
408 *hns* double mutant (Fig. 4B). The amount of Vat secreted by CFT073*hns* and CFT073*vatX hns*
409 was markedly increased compared to WT CFT073. Consistent with previous results, the level of
410 Vat detected in supernatant fraction of CFT073*vatX* was similar to that detected from WT
411 CFT073.
412
413 **H-NS binds to the *vat* promoter region.** To further investigate the role of H-NS in repression of
414 *vat* transcription, an EMSA was performed using increasing concentrations of native H-NS
415 protein and the 252bp PCR amplified region of the *vat* gene promoter possessing the three
416 potential H-NS binding sites (Fig 4C). As a positive control, the *bla* gene promoter from the
417 cloning vector pBR322 was also PCR amplified and included in the assay; H-NS is known to
418 bind to this DNA fragment (91). Three additional fragments amplified from regions of pBR322
419 known not to bind H-NS were included in the assay as negative controls. In our experiment, H-
420 NS bound with strong affinity to the DNA fragment corresponding to the *vat* gene promoter.
421 Indeed, this binding affinity was stronger than that observed for the DNA fragment containing
422 the control *bla* gene promoter. No binding of H-NS was observed to the negative control DNA
423 fragments, demonstrating the specificity of H-NS binding in this assay.
424
425 **The *vatX* gene is co-transcribed with *vat*.** H-NS regulates the transcription of several UPEC
426 genes by competing for binding to their promoter element with a MarR-type regulatory protein;
427 this includes SfaX binding to the sfa_{II} fimbrial promoter (80), PapX binding to the *flhDC* flagella
428 master regulator promoter (92), and SlyA binding to the type 1 fimbriae *fimB* recombinase
429 promoter (93). The SfaX and PapX regulators are co-transcribed as part of their respective
430 upstream fimbrial operon (encoding S and P type fimbriae, respectively (47, 80)). Taking this

into consideration, we employed RT-PCR analysis to test for transcription of the *vat* and *vatX* genes as a single mRNA in CFT073. Due to the increased amount of Vat protein secreted by the CFT073*hns* mutant strain (as shown by Western blot analysis), total RNA was extracted from this strain, converted to cDNA and screened for a *vat-vatX* nucleic acid fragment using internal primers specific for both genes by RT-PCR (Fig. 4D). For comparison, an additional set of primers were used to amplify the *vatX* gene alone. Bands corresponding to the predicted sizes determined for the *vatX* and the *vat-vatX* transcripts were amplified from CFT073*hns* cDNA. Thus, while we cannot rule out the presence of an independent promoter upstream of *vatX*, our results demonstrate that the *vat-vatX* genes are co-transcribed in the absence of H-NS.

Vat is prevalent, highly conserved and is secreted by UPEC urosepsis isolates. The *vat* gene has previously been shown to be most prevalent in *E. coli* strains from the B2 phylogenetic group, with a similar distribution observed among cystitis, pyelonephritis, prostatitis and bloodstream isolates (45). Based on the observation that *vat* is required for UPEC fitness in a mouse model of systemic infection (50), we screened a collection of urosepsis strains for the *vat* gene by PCR. The *vat* gene was identified in 68% (31/45) of the urosepsis strains. MLST analysis revealed strains from ST73, ST12 and ST95 were most predominant in this collection (Fig. 5). Furthermore, supernatant fractions produced by these strains were examined by Western blotting to analyse the expression and secretion of Vat following growth in LB at 37°C. For all strains, a band corresponding to the Vat passenger domain hybridised with the Vat-specific polyclonal antibody. The sequence of the *vat* gene was determined from *vat*-positive strains representing each ST and found to be highly conserved (≥97% amino acid identity [Fig. 5]). Minor sequence variations occurred at six locations within the passenger domain of the protein. These residues were located within two regions in the Vat passenger domain (Fig. 5), both of which are distal to the serine protease catalytic motif based on a structural model built using the Hbp passenger domain (Fig. S2).

The presence of *vat* is associated with increased anti-Vat IgG produced during infection. The high prevalence of *vat* in the UPEC urosepsis strains examined in this study, in combination with its secretion during *in vitro* growth, prompted us to examine if an immunological response against Vat was elicited during infection. To test this, an ELISA assay was performed using

blood plasma samples collected from the same urosepsis patients from which the urosepsis strains were collected (Fig. 6). The blood plasma samples were examined for the presence of Vat-specific IgG antibodies using purified recombinant Vat protein. The urosepsis patients were divided into two groups; those infected with a *vat*-positive UPEC strain (n=31) and those infected with a *vat*-negative UPEC strain (n=14). As an additional control, 42 plasma samples collected from age and sex matched healthy individuals were also examined for an immunological response against the Vat protein. In this assay, we observed a significant difference ($P < 0.05$) in the anti-Vat IgG plasma titre in patients infected with a *vat*-positive strain compared to a *vat*-negative strain or healthy individuals. Taken together, these data suggest that Vat is a highly conserved immunogenic protein that is expressed by many UPEC isolates during infection.

DISCUSSION

UPEC strains possess an array of virulence factors that are critical for their ability to cause disease in extra-intestinal niches such as the urinary tract and the bloodstream (94, 95). Vat is a member of the SPATEs that contributes to fitness of *E. coli* during systemic infection (46, 50). In this study, we performed a comprehensive bioinformatic and molecular analysis of the *vat* gene. We defined the transcriptional regulation of *vat* and demonstrated its immunogenicity using plasma samples from urosepsis patients.

The genomic location of the *vat* gene was examined in all *vat*-positive completely sequenced *E. coli* strains available on the NCBI database. The *vat* gene was shown to reside within the *thrW*-PAI, downstream of *proA* and upstream of *yagU* relative to the *E. coli* MG1655 chromosome. This is consistent with a previous report that examined the presence of *vat* in UPEC strains CFT073 and 536, as well as the neonatal meningitis strain RS218 (45). The gene content of the *vat*-containing *thrW*-PAI was conserved in the majority of strains examined, although some differences were noted in strains Ec222, APEC-O1, 83972, UM146 and 536. Overall, our bioinformatic analysis revealed that the *vat* gene (and the co-located *vatX* regulator gene) is present in a range of different *E. coli* pathoypes.

493 Several studies have previously assessed the prevalence of the *vat* gene in *E. coli*. A study
494 conducted by Parham *et al* (45) reported a high prevalence of *vat* in group B2 phylogenetic
495 strains of the ECOR collection. A high frequency of the *vat* gene has also been observed in B2
496 strains associated with cystitis, pyelonephritis and prostatitis (45, 59), and *vat* has been strongly
497 associated with avian pathogenic *E. coli* (APEC) (96). Our analysis identified the *vat* gene in
498 68% of urosepsis isolates (n = 45). We also demonstrated that the sequence of *vat* is highly
499 conserved within a selection of strains representative of each of the ten different sequence types
500 identified in our collection. At the amino acid level, minor sequence variations were located
501 within two regions (VR1: S^{520}-K^{529} and VR2: E^{783}-V^{823}) of the Vat passenger domain. However,
502 the canonical serine protease domain that is important for the catalytic function of SPATEs was
503 conserved in all ten of the Vat sequences analysed. Western blotting was also performed to
504 examine Vat expression, and revealed that Vat is expressed and secreted by all of the urosepsis
505 strains examined when grown at human core body temperature. Further investigation is required
506 to determine whether the minor sequence changes observed in Vat are associated with
507 corresponding differences in its cytotoxic properties.
508
509 Bioinformatic analysis identified a gene encoding a putative MarR-like transcriptional regulator
510 immediately downstream of *vat* (i.e. *vatX*). Although mutation of *vatX* did not result in a
511 detectable change in *vat* transcription or translation, overexpression of VatX via the introduction
512 of a plasmid containing the *vatX* gene (pVatX) was shown to positively regulate *vat*, resulting in
513 a 3-fold increase in *vat* transcription and a significant increase in the level of secreted Vat
514 protein. This data was suggestive of a more complex regulatory control of the *vat* gene. We
515 therefore mapped the promoter of *vat*, and identified several putative H-NS binding sites
516 proximal to this region. H-NS is a histone-like DNA-binding protein that shows affinity for A-T
517 rich and bent nucleation sites on DNA (97). In *E. coli*, H-NS has been shown to regulate multiple
518 genes, including genes associated with virulence, pH, osmoregulation and temperature sensing
519 (98-101). Our EMSA data revealed a strong interaction between H-NS and a 252-bp region of
520 the *vat* promoter that contains three putative H-NS binding sites. A role for H-NS in *vat*
521 regulation was subsequently demonstrated through the examination of a CFT073*hns* mutant,
522 which secreted a significantly higher level of Vat compared to the parent CFT073 strain. Taken

together, these results demonstrate that the regulation of *vat* is coordinated by both VatX and H-NS, and further highlights the role of H-NS in the regulation of UPEC virulence factors (8, 9).

The MarR family of transcriptional regulators control the expression of multiple different genes, including virulence factors, often in response to environmental stress (reviewed in 102, 103). Bioinformatic analysis of MarR-type regulators from 77 completely sequenced *E. coli* genomes revealed a high level of amino acid sequence conservation for proteins in each clade, but significant variation between MarR regulators from different clades. VatX clustered as a separate clade and is most closely related to PapX. Interestingly, the proteins encoding for other fimbrial associated MarR-type regulators were also found within the PapX clade (Fig S1). Despite their association with different fimbriae, these regulatory proteins are highly conserved (≥97% amino acid identity). Some strains such as *E. coli* 536, 83972 and Nissle 1917 possess three or more chromosomal copies of *papX* (Table S2). PapX regulates UPEC motility by repressing transcription of the *flhDC* master regulator genes (47). We investigated the potential for VatX to repress flagella-mediated motility of CFT073. However, no significant difference in motility was observed between WT CFT073, CFT073*vatX* and the complemented CFT073*vatX* (pVatX) mutant strains after growth at 28°C and 37°C (data not shown). The FliC major flagellin subunit was also produced at a similar level in all three strains as determined by immunoblotting (data not shown). Taken together, our data has identified VatX as a new member of the MarR type family that appears to regulate *vat* in concert with H-NS. Further work is now required to map the direct binding of VatX to the *vat* gene promoter, and to examine the competitive interplay between VatX and H-NS in the regulation of *vat* transcription.

In a recent study using high-throughput transposon mutagenesis screening (50), the *vat* gene was shown to contribute to survival of the UPEC strain CFT073 in the bloodstream of mice. This, together with the observation that many urosepsis strains secrete Vat, prompted us to examine the immunoreactivity of Vat in urosepsis patients. We detected a significant increase in Vat-specific IgG titre in the plasma of urosepsis patients infected with *vat*-positive UPEC strains compared to plasma from patients infected with *vat*-negative strains and healthy controls. Although we cannot rule out that the responses we detected may in part be due to previous or ongoing infection that culminated in sepsis, overall the data is consistent with the notion that Vat

554 is expressed during infection and elicits a strong immune response in some patients. Further
555 work is now required to understand the role of Vat during human infection and its cytotoxicity
556 profile.
557

558 ## ACKNOWLEDGEMENTS

559 We thank David Looke, Joan Faoagali and other members of the Microbiology Lab, Princess
560 Alexandra Hospital, for the collection of urosepsis strains and plasma samples, and Barbara
561 Johnson for the collection of patient clinical data. This work was supported by a grant from the
562 National Health and Medical Research Council (NHMRC) of Australia (APP1042651). MAS is
563 supported by an NHMRC Senior Research Fellowship (APP1106930), SAB by an NHMRC
564 Career Development Fellowship (APP1090456), GCU by an Australian Research Council
565 (ARC) Future Fellowship (FT110101048) and MT by an ARC Discovery Early Career
566 Researcher Award (DE130101169).

REFERENCES

1. **Foxman B.** 2010. The epidemiology of urinary tract infection. Nat Rev Urol **7:**653-660.
2. **Hooton TM, Stamm WE.** 1997. Diagnosis and treatment of uncomplicated urinary tract infection. Infect Dis Clin North Am **11:**551-581.
3. **Connell H, Agace W, Klemm P, Schembri M, Marild S, Svanborg C.** 1996. Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. Proc Natl Acad Sci U S A **93:**9827-9832.
4. **Mulvey MA, Lopez-Boado YS, Wilson CL, Roth R, Parks WC, Heuser J, Hultgren SJ.** 1998. Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. Science **282:**1494-1497.
5. **Wu XR, Sun TT, Medina JJ.** 1996. *In vitro* binding of type 1-fimbriated *Escherichia coli* to uroplakins Ia and Ib: relation to urinary tract infections. Proc Natl Acad Sci U S A **93:**9630-9635.
6. **Roberts JA, Marklund BI, Ilver D, Haslam D, Kaack MB, Baskin G, Louis M, Mollby R, Winberg J, Normark S.** 1994. The Gal(alpha 1-4)Gal-specific tip adhesin of *Escherichia coli* P-fimbriae is needed for pyelonephritis to occur in the normal urinary tract. Proc Natl Acad Sci U S A **91:**11889-11893.
7. **Ulett GC, Valle J, Beloin C, Sherlock O, Ghigo JM, Schembri MA.** 2007. Functional analysis of antigen 43 in uropathogenic *Escherichia coli* reveals a role in long-term persistence in the urinary tract. Infect Immun **75:**3233-3244.
8. **Allsopp LP, Beloin C, Ulett GC, Valle J, Totsika M, Sherlock O, Ghigo JM, Schembri MA.** 2012. Molecular Characterization of UpaB and UpaC, Two New Autotransporter Proteins of Uropathogenic *Escherichia coli* CFT073. Infect Immun **80:**321-332.
9. **Allsopp LP, Totsika M, Tree JJ, Ulett GC, Mabbett AN, Wells TJ, Kobe B, Beatson SA, Schembri MA.** 2010. UpaH Is a Newly Identified Autotransporter Protein That Contributes to Biofilm Formation and Bladder Colonization by Uropathogenic *Escherichia coli* CFT073. Infect Immun **78:**1659-1669.
10. **Totsika M, Wells TJ, Beloin C, Valle J, Allsopp LP, King NP, Ghigo JM, Schembri MA.** 2012. Molecular Characterization of the EhaG and UpaG Trimeric Autotransporter Proteins from Pathogenic *Escherichia coli*. Appl Environ Microbiol **78:**2179-2189.
11. **Anderson GG, Goller CC, Justice S, Hultgren SJ, Seed PC.** 2010. Polysaccharide capsule and sialic acid-mediated regulation promote biofilm-like intracellular bacterial communities during cystitis. Infect Immun **78:**963-975.
12. **Sarkar S, Ulett GC, Totsika M, Phan MD, Schembri MA.** 2014. Role of capsule and O antigen in the virulence of uropathogenic *Escherichia coli*. PLoS One **9:**e94786.
13. **Phan MD, Peters KM, Sarkar S, Lukowski SW, Allsopp LP, Gomes Moriel D, Achard ME, Totsika M, Marshall VM, Upton M, Beatson SA, Schembri MA.** 2013. The serum resistome of a globally disseminated multidrug resistant uropathogenic *Escherichia coli* clone. PLoS Genet **9:**e1003834.
14. **Valdebenito M, Bister B, Reissbrodt R, Hantke K, Winkelmann G.** 2005. The detection of salmochelin and yersiniabactin in uropathogenic *Escherichia coli* strains by a novel hydrolysis-fluorescence-detection (HFD) method. Int J Med Microbiol **295:**99-107.
15. **Garcia EC, Brumbaugh AR, Mobley HLT.** 2011. Redundancy and Specificity of *Escherichia coli* Iron Acquisition Systems during Urinary Tract Infection. Infect Immun **79:**1225-1235.

16. **Watts RE, Totsika M, Challinor VL, Mabbett AN, Ulett GC, De Voss JJ, Schembri MA.** 2012. Contribution of siderophore systems to growth and urinary tract colonization of asymptomatic bacteriuria *Escherichia coli*. Infect Immun **80:**333-344.
17. **Guyer DM, Radulovic S, Jones FE, Mobley HL.** 2002. Sat, the secreted autotransporter toxin of uropathogenic *Escherichia coli*, is a vacuolating cytotoxin for bladder and kidney epithelial cells. Infect Immun **70:**4539-4546.
18. **Rippere-Lampe KE, O'Brien AD, Conran R, Lockman HA.** 2001. Mutation of the gene encoding cytotoxic necrotizing factor type 1 (cnf(1)) attenuates the virulence of uropathogenic *Escherichia coli*. Infect Immun **69:**3954-3964.
19. **Dhakal Bijaya K, Mulvey Matthew A.** 2012. The UPEC Pore-Forming Toxin α-Hemolysin Triggers Proteolysis of Host Proteins to Disrupt Cell Adhesion, Inflammatory, and Survival Pathways. Cell Host & Microbe **11:**58-69.
20. **Garcia TA, Ventura CL, Smith MA, Merrell DS, O'Brien AD.** 2013. Cytotoxic necrotizing factor 1 and hemolysin from uropathogenic *Escherichia coli* elicit different host responses in the murine bladder. Infect Immun **81:**99-109.
21. **Smith YC, Rasmussen SB, Grande KK, Conran RM, O'Brien AD.** 2008. Hemolysin of uropathogenic *Escherichia coli* evokes extensive shedding of the uroepithelium and hemorrhage in bladder tissue within the first 24 hours after intraurethral inoculation of mice. Infect Immun **76:**2978-2990.
22. **Leyton DL, Rossiter AE, Henderson IR.** 2012. From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis. Nat Rev Microbiol **10:**213-225.
23. **Leo JC, Grin I, Linke D.** 2012. Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. Philos Trans R Soc Lond B Biol Sci **367:**1088-1101.
24. **Gawarzewski I, Smits SH, Schmitt L, Jose J.** 2013. Structural comparison of the transport units of type V secretion systems. Biol Chem **394:**1385-1398.
25. **Grijpstra J, Arenas J, Rutten L, Tommassen J.** 2013. Autotransporter secretion: varying on a theme. Res Microbiol **164:**562-582.
26. **Benz I, Schmidt MA.** 2011. Structures and functions of autotransporter proteins in microbial pathogens. Int J Med Microbiol **301:**461-468.
27. **Ieva R, Bernstein HD.** 2009. Interaction of an autotransporter passenger domain with BamA during its translocation across the bacterial outer membrane. Proc Natl Acad Sci U S A **106:**19120-19125.
28. **Sauri A, Soprova Z, Wickstrom D, de Gier JW, Van der Schors RC, Smit AB, Jong WS, Luirink J.** 2009. The Bam (Omp85) complex is involved in secretion of the autotransporter haemoglobin protease. Microbiology **155:**3982-3991.
29. **Selkrig J, Mosbahi K, Webb CT, Belousoff MJ, Perry AJ, Wells TJ, Morris F, Leyton DL, Totsika M, Phan MD, Celik N, Kelly M, Oates C, Hartland EL, Robins-Browne RM, Ramarathinam SH, Purcell AW, Schembri MA, Strugnell RA, Henderson IR, Walker D, Lithgow T.** 2012. Discovery of an archetypal protein transport system in bacterial outer membranes. Nat Struct Mol Biol **19:**506-510, S501.
30. **Jain S, Goldberg MB.** 2007. Requirement for YaeT in the outer membrane assembly of autotransporter proteins. J Bacteriol **189:**5393-5398.

31. **Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC, Ala'Aldeen D.** 2004. Type V protein secretion pathway: the autotransporter story. Microbiol Mol Biol Rev **68:**692-744.
32. **Desvaux M, Parham NJ, Henderson IR.** 2004. Type V protein secretion: simplicity gone awry? Curr Issues Mol Biol **6:**111-124.
33. **Nishimura K, Tajima N, Yoon YH, Park SY, Tame JR.** 2010. Autotransporter passenger proteins: virulence factors with common structural themes. J Mol Med (Berl) **88:**451-458.
34. **Yen YT, Kostakioti M, Henderson IR, Stathopoulos C.** 2008. Common themes and variations in serine protease autotransporters. Trends Microbiol **16:**370-379.
35. **Dautin N.** 2010. Serine protease autotransporters of *Enterobacteriaceae* (SPATEs): biogenesis and function. Toxins (Basel) **2:**1179-1206.
36. **Ruiz-Perez F, Nataro JP.** 2014. Bacterial serine proteases secreted by the autotransporter pathway: classification, specificity, and role in virulence. Cell Mol Life Sci **71:**745-770.
37. **Dutta PR, Cappello R, Navarro-Garcia F, Nataro JP.** 2002. Functional comparison of serine protease autotransporters of *Enterobacteriaceae*. Infect Immun **70:**7105-7113.
38. **Djafari S, Ebel F, Deibel C, Kramer S, Hudel M, Chakraborty T.** 1997. Characterization of an exported protease from Shiga toxin-producing *Escherichia coli*. Mol Microbiol **25:**771-784.
39. **Stein M, Kenny B, Stein MA, Finlay BB.** 1996. Characterization of EspC, a 110-kilodalton protein secreted by enteropathogenic *Escherichia coli* which is homologous to members of the immunoglobulin A protease-like family of secreted proteins. J Bacteriol **178:**6546-6554.
40. **Guyer DM, Henderson IR, Nataro JP, Mobley HL.** 2000. Identification of Sat, an autotransporter toxin produced by uropathogenic *Escherichia coli*. Mol Microbiol **38:**53-66.
41. **Al-Hasani K, Henderson IR, Sakellaris H, Rajakumar K, Grant T, Nataro JP, Robins-Browne R, Adler B.** 2000. The *sigA* gene which is borne on the *she* pathogenicity island of *Shigella flexneri 2a* encodes an exported cytopathic protease involved in intestinal fluid accumulation. Infect Immun **68:**2457-2463.
42. **Navarro-Garcia F, Canizalez-Roman A, Sui BQ, Nataro JP, Azamar Y.** 2004. The serine protease motif of EspC from enteropathogenic *Escherichia coli* produces epithelial damage by a mechanism different from that of pet toxin from enteroaggregative *E. coli*. Infect Immun **72:**3609-3621.
43. **Henderson IR, Hicks S, Navarro-Garcia F, Elias WP, Philips AD, Nataro JP.** 1999. Involvement of the enteroaggregative *Escherichia coli* plasmid-encoded toxin in causing human intestinal damage. Infect Immun **67:**5338-5344.
44. **Ruiz-Perez F, Wahid R, Faherty CS, Kolappaswamy K, Rodriguez L, Santiago A, Murphy E, Cross A, Sztein MB, Nataro JP.** 2011. Serine protease autotransporters from *Shigella flexneri* and pathogenic *Escherichia coli* target a broad range of leukocyte glycoproteins. Proc Natl Acad Sci U S A **108:**12881-12886.
45. **Parham NJ, Pollard SJ, Desvaux M, Scott-Tucker A, Liu C, Fivian A, Henderson IR.** 2005. Distribution of the serine protease autotransporters of the *Enterobacteriaceae* among extraintestinal clinical isolates of *Escherichia coli*. J Clin Microbiol **43:**4076-4082.

46. **Parreira VR, Gyles CL.** 2003. A novel pathogenicity island integrated adjacent to the *thrW* tRNA gene of avian pathogenic *Escherichia coli* encodes a vacuolating autotransporter toxin. Infect Immun **71:**5087-5096.
47. **Simms AN, Mobley HL.** 2008. PapX, a P fimbrial operon-encoded inhibitor of motility in uropathogenic *Escherichia coli*. Infect Immun **76:**4833-4841.
48. **Reiss DJ, Mobley HL.** 2011. Determination of target sequence bound by PapX, repressor of bacterial motility, in *flhD* promoter using systematic evolution of ligands by exponential enrichment (SELEX) and high throughput sequencing. J Biol Chem **286:**44726-44738.
49. **Simms AN, Mobley HL.** 2008. Multiple genes repress motility in uropathogenic *Escherichia coli* constitutively expressing type 1 fimbriae. J Bacteriol **190:**3747-3756.
50. **Subashchandrabose S, Smith SN, Spurbeck RR, Kole MM, Mobley HLT.** 2013. Genome-Wide Detection of Fitness Genes in Uropathogenic *Escherichia coli* during Systemic Infection. PLoS Pathog **9**.
51. **Otto BR, van Dooren SJ, Nuijens JH, Luirink J, Oudega B.** 1998. Characterization of a hemoglobin protease secreted by the pathogenic *Escherichia coli* strain EB1. J Exp Med **188:**1091-1103.
52. **Provence DL, Curtiss R, 3rd.** 1994. Isolation and characterization of a gene involved in hemagglutination by an avian pathogenic *Escherichia coli* strain. Infect Immun **62:**1369-1380.
53. **Otto BR, Sijbrandi R, Luirink J, Oudega B, Heddle JG, Mizutani K, Park SY, Tame JR.** 2005. Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic *Escherichia coli*. J Biol Chem **280:**17339-17345.
54. **Tame JRH, van Dooren SJM, Oudega B, Otto BR.** 2002. Characterization and crystallization of a novel haemoglobinase from pathogenic *Escherichia coli*. Acta Crystallogr D Biol Crystallogr **58:**843-845.
55. **Nishimura K, Yoon YH, Kurihara A, Unzai S, Luirink J, Park SY, Tame JRH.** 2010. Role of domains within the autotransporter Hbp/Tsh. Acta Crystallogr D Biol Crystallogr **66:**1295-1300.
56. **van Dooren SJM, Tame JRH, Luirink J, Oudega B, Otto BR.** 2001. Purification of the autotransporter protein Hbp of *Escherichia coli*. Fems Microbiol Lett **205:**147-150.
57. **Kobayashi RK, Gaziri LC, Vidotto MC.** 2010. Functional activities of the Tsh protein from avian pathogenic *Escherichia coli* (APEC) strains. J Vet Sci **11:**315-319.
58. **Welch RA, Burland V, Plunkett G, 3rd, Redford P, Roesch P, Rasko D, Buckles EL, Liou SR, Boutin A, Hackett J, Stroud D, Mayhew GF, Rose DJ, Zhou S, Schwartz DC, Perna NT, Mobley HL, Donnenberg MS, Blattner FR.** 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc Natl Acad Sci U S A **99:**17020-17024.
59. **Heimer SR, Rasko DA, Lockatell CV, Johnson DE, Mobley HLT.** 2004. Autotransporter genes *pic* and *tsh* are associated with *Escherichia coli* strains that cause acute pyelonephritis and are expressed during urinary tract infection. Infect Immun **72:**593-597.
60. **Mobley HLT, Green DM, Trifillis AL, Johnson DE, Chippendale GR, Lockatell CV, Jones BD, Warren JW.** 1990. Pyelonephritogenic *Escherichia coli* and Killing of Cultured Human Renal Proximal Tubular Epithelial Cells: Role of Hemolysin in Some Strains. Infect Immun **58:**1281-1289.

61. **Moriel DG, Bertoldi I, Spagnuolo A, Marchi S, Rosini R, Nesta B, Pastorello I, Corea VA, Torricelli G, Cartocci E, Savino S, Scarselli M, Dobrindt U, Hacker J, Tettelin H, Tallon LJ, Sullivan S, Wieler LH, Ewers C, Pickard D, Dougan G, Fontana MR, Rappuoli R, Pizza M, Serino L.** 2010. Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic Escherichia coli. Proc Natl Acad Sci U S A **107:**9072-9077.
62. **LeClerc JE, Li BG, Payne WL, Cebula TA.** 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. Science **274:**1208-1211.
63. **Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453-1462.
64. **Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, Hur CG, Park HS, Segurens B, Kim SC, Oh TK, Lenski RE, Studier FW, Daegelen P, Kim JF.** 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). J Mol Biol **394:**644-652.
65. **Ochman H, Selander RK.** 1984. Standard reference strains of *Escherichia coli* from natural populations. J Bacteriol **157:**690-693.
66. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944-945.
67. **Sullivan MJ, Petty NK, Beatson SA.** 2011. Easyfig: a genome comparison visualizer. Bioinformatics **27:**1009-1010.
68. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG.** 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol **7:**539.
69. **Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R.** 2010. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res **38:**W695-699.
70. **Rambaut A, Drummond A.** 2009. FigTree, v1.4.2. http://tree.bio.ed.ac.uk/software/figtree/.
71. **Sambrook J, Fritsch, E. F. & Maniatis, T.** 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, New York.
72. **Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M.** 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol **60:**1136-1151.
73. **Datsenko KA, Wanner BL.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A **97:**6640-6645.
74. **Martinez E, Bartolome B, de la Cruz F.** 1988. pACYC184-derived cloning vectors containing the multiple cloning site and *lacZ*α reporter gene of pUC8/9 and pUC18/19 plasmids. Gene **68:**159-162.
75. **Livak KJ, Schmittgen TD.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods **25:**402-408.
76. **Munch R, Hiller K, Grote A, Scheer M, Klein J, Schobert M, Jahn D.** 2005. Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. Bioinformatics **21:**4187-4189.

77. **Beloin C, Dorman CJ.** 2003. An extended role for the nucleoid structuring protein H-NS in the virulence gene regulatory cascade of *Shigella flexneri*. Mol Microbiol **47:**825-838.
78. **Ulett GC, Webb RI, Schembri MA.** 2006. Antigen-43-mediated autoaggregation impairs motility in *Escherichia coli*. Microbiology **152:**2101-2110.
79. **Delcastillo I, Gomez JM, Moreno F.** 1990. *mprA*, an *Escherichia coli* Gene That Reduces Growth-Phase-Dependent Synthesis of Microcin-B17 and Microcin-C7 and Blocks Osmoinduction of *proU* When Cloned on a High-Copy-Number Plasmid. J Bacteriol **172:**437-445.
80. **Sjostrom AE, Sonden B, Muller C, Rydstrom A, Dobrindt U, Wai SN, Uhlin BE.** 2009. Analysis of the *sfaX(II)* locus in the *Escherichia coli* meningitis isolate IHE3034 reveals two novel regulatory genes within the promoter-distal region of the main S fimbrial operon. Microb Pathog **46:**150-158.
81. **Lomovskaya O, Lewis K, Matin A.** 1995. Emrr Is a Negative Regulator of the *Escherichia coli* Multidrug Resistance Pump Emrab. J Bacteriol **177:**2328-2334.
82. **Roper DI, Fawcett T, Cooper RA.** 1993. The *Escherichia coli* C Homoprotocatechuate Degradative Operon: *hpc* Gene Order, Direction of Transcription and Control of Expression. Mol Gen Genet **237:**241-250.
83. **Paracuellos P, Ohman A, Sauer-Eriksson AE, Uhlin BE.** 2012. Expression and purification of SfaX(II), a protein involved in regulating adhesion and motility genes in extraintestinal pathogenic *Escherichia coli*. Protein Expr Purif **86:**127-134.
84. **Dezfulian H, Tremblay D, Harel J.** 2004. Molecular characterization of extraintestinal pathogenic *Escherichia coli* (ExPEC) pathogenicity islands in F165-positive *E. coli* strain from a diseased animal. FEMS Microbiol Lett **238:**321-332.
85. **Lang B, Blot N, Bouffartigues E, Buckle M, Geertz M, Gualerzi CO, Mavathur R, Muskhelishvili G, Pon CL, Rimsky S, Stella S, Babu MM, Travers A.** 2007. High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. Nucleic Acids Research **35:**6330-6337.
86. **Yang J, Tauschek M, Strugnell R, Robins-Browne RM.** 2005. The H-NS protein represses transcription of the *eltAB* operon, which encodes heat-labile enterotoxin in enterotoxigenic *Escherichia coli*, by binding to regions downstream of the promoter. Microbiology-Sgm **151:**1199-1208.
87. **Scott ME, Melton-Celsa AR, O'Brien AD.** 2003. Mutations in *hns* reduce the adherence of Shiga toxin-producing *E. coli* 091 : H21 strain B2F1 to human colonic epithelial cells and increase the production of hemolysin. Microb Pathog **34:**155-159.
88. **Lithgow JK, Haider F, Roberts IS, Green J.** 2007. Alternate SlyA and H-NS nucleoprotein complexes control *hlyE* expression in *Escherichia coli* K-12. Mol Microbiol **66:**685-698.
89. **Li HP, Granat A, Stewart V, Gillespie JR.** 2008. RpoS, H-NS, and DsrA influence EHEC hemolysin operon (*ehxCABD*) transcription in *Escherichia coli* O157 : H7 strain EDL933. FEMS Microbiol Lett **285:**257-262.
90. **Allsopp LP, Beloin C, Moriel DG, Totsika M, Ghigo JM, Schembri MA.** 2012. Functional heterogeneity of the UpaH autotransporter protein from uropathogenic *Escherichia coli*. J Bacteriol **194:**5769-5782.
91. **Bolivar F, Rodriguez RL, Greene PJ, Betlach MC, Heyneker HL, Boyer HW, Crosa JH, Falkow S.** 1977. Construction and characterization of new cloning vehicle. II. A multipurpose cloning system. Gene **2:**95-113.

92. **Ko M, Park C.** 2000. Two novel flagellar components and H-NS are involved in the motor function of *Escherichia coli*. J Mol Biol **303:**371-382.
93. **McVicker G, Sun L, Sohanpal BK, Gashi K, Williamson RA, Plumbridge J, Blomfield IC.** 2011. SlyA Protein Activates *fimB* Gene Expression and Type 1 Fimbriation in *Escherichia coli* K-12. Journal of Biological Chemistry **286:**32026-32035.
94. **Nielubowicz GR, Mobley HL.** 2010. Host-pathogen interactions in urinary tract infection. Nat Rev Urol **7:**430-441.
95. **Totsika M, Moriel DG, Idris A, Rogers BA, Wurpel DJ, Phan MD, Paterson DL, Schembri MA.** 2012. Uropathogenic *Escherichia coli* mediated urinary tract infection. Curr Drug Targets **13:**1386-1399.
96. **Spurbeck RR, Dinh PC, Walk ST, Stapleton AE, Hooton TM, Nolan LK, Kim KS, Johnson JR, Mobley HLT.** 2012. *Escherichia coli* Isolates That Carry *vat*, *fyuA*, *chuA*, and *yfcV* Efficiently Colonize the Urinary Tract. Infect Immun **80:**4115-4122.
97. **Dorman CJ.** 2007. H-NS, the genome sentinel. Nat Rev Microbiol **5:**157-161.
98. **Goransson M, Sonden B, Nilsson P, Dagberg B, Forsman K, Emanuelsson K, Uhlin BE.** 1990. Transcriptional silencing and thermoregulation of gene expression in *Escherichia coli*. Nature **344:**682-685.
99. **Atlung T, Ingmer H.** 1997. H-NS: a modulator of environmentally regulated gene expression. Mol Microbiol **24:**7-17.
100. **Amit R, Oppenheim AB, Stavans J.** 2003. Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. Biophys J **84:**2467-2473.
101. **Lucchini S, Rowley G, Goldberg MD, Hurd D, Harrison M, Hinton JC.** 2006. H-NS mediates the silencing of laterally acquired genes in bacteria. PLoS Pathog **2:**e81.
102. **Wilkinson SP, Grove A.** 2006. Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. Curr Issues Mol Biol **8:**51-62.
103. **Ellison DW, Miller VL.** 2006. Regulation of virulence by members of the MarR/SlyA family. Curr Opin Microbiol **9:**153-159.

FIGURE LEGENDS

FIG. 1. (A) BLAST alignment demonstrating the level of nucleotide sequence conservation (grey shading) for *vat* and *vatX* (labelled red), as well as the other surrounding genes (labelled blue). The Vat-PAI (defined by the *proA* and *yagU* genes [labelled yellow]) was identified in 14 of 77 complete *E. coli* genomes examined. These sequences were compared to the Vat-PAI originally identified in the avian pathogenic *E. coli* strain Ec222 (top). (B) Immunodetection of the Vat passenger domain (Vatα) from supernatant fractions prepared from overnight cultures of the well-characterised UPEC strains CFT073, IHE3034 and 536. Vat expression by MG1655 (pVat) is shown as a positive control, while MG1655(pSU2718) and CFT073*vat* were included as a negative controls.

881 **FIG. 2.** Phylogram demonstrating the relationship between representative *E. coli* MarR-type
882 regulator proteins. The scale represents the number of amino acid substitutions per site over 194
883 positions.
884
885 **FIG. 3.** (A) qRT-PCR analysis of *vat* transcription in CFT073*vatX* and CFT073*vatX*(pVatX)
886 compared to wild-type CFT073. The transcription of *vat* was significantly increased in
887 CFT073*vatX*(pVatX) compared to CFT073 (**P<0.01). (B) Western blot analysing the effect of
888 VatX on Vat expression. Supernatant fractions were prepared from overnight cultures of
889 MG1655(pVat), MG1655(pSU2718), CFT073(pSU2718), CFT073(pVatX),
890 CFT073*vatX*(pSU2718) and CFT073*vatX*(pVatX). Over-expression of VatX led to an increase in
891 the amount Vat detected in the culture supernatant.
892
893 **FIG. 4.** (A) Schematic of the *vat-vatX* gene operon in CFT073. The position of the promoter and
894 primers used to identify *vat-vatX* and *vatX* transcripts is indicated. The inset shows the *vat* gene
895 transcriptional start site (+1), which was mapped to 80bp upstream of the ATG start codon (grey
896 arrow). Also indicated are the consensus -10 and -35 promoter elements and the three putative H-
897 NS nucleation sites (shown in bold). (B) Immunodetection of the Vat passenger domain from the
898 supernatant fractions of CFT073 , CFT073*vat*, CFT073*vatX*, CFT073*hns* and CFT073*vatX hns*.
899 The level of Vat was increased in CFT073*hns* and CFT073*vatX hns* compared to CFT073. (C)
900 EMSA demonstrating the direct interaction of H-NS with the *vat* promoter region. The assay was
901 performed using a 252bp fragment encompassing the *vat* promoter region (indicated by an
902 asterisk), a 218bp fragment containing the *bla* promoter region amplified from pBR322 (positive
903 control: indicated by an arrow), and three additional DNA fragments amplified from pBR322
904 (negative controls: 152bp, 312bp and 479bp). Native H-NS protein was incubated with the DNA
905 in increasing concentrations (0μM H-NS, 0.1μM H-NS, 0.5μM H-NS and 1.0μM H-NS). (D)
906 Transcriptional analysis of the *vat* and *vatX* genes. Total RNA was extracted during exponential
907 growth of CFT073*hns* and converted to cDNA. Shown are the PCR products [*vat-vatX* (1112bp)
908 or *vatX* (404bp)] amplified from CFT073*hns* gDNA (positive control), total RNA (negative
909 control) and cDNA.
910

FIG. 5. (A) Diagram depicting the full length Vat primary protein sequence, including three protein domains typical for SPATES: i) the extended signal peptide (SP); ii) the passenger domain comprising the Immunoglobulin A1 protease-like domain, which contains the serine protease motif, as well as the upstream aspartate (D158) and histidine (H130) residues of the catalytic triad; and iii) the translocation domain, which is cleaved at the alpha-helical linker region. Class II SPATEs are characterised by the presence of a small additional domain termed Domain 2 (striped). Two variable regions (VR1 and VR2) located within the passenger domain were identified (triangles). (B) Alignment of the Vat amino acid sequence in VR1 and VR2 from CFT073 and the ten strains representing the diverse STs examined. Residues identical to those in Vat from CFT073 are indicated by dots; residues that differed from the CFT073 sequence are indicated and highlighted in grey. Vat secretion was determined by Western blot analysis of the supernatant fractions from each strain following overnight growth in LB broth at 37°C. All strains secreted a ~110kDa protein that cross-reacted with the Vat-specific polyclonal antibody (indicated as +).

FIG. 6. Immunoreactivity of Vat. Blood plasma was collected from 45 urosepsis patients at the time of admittance to hospital. Paired UPEC strains were also isolated from the blood of each patient, and the presence of the *vat* gene was determined by PCR. Plasma samples were subsequently grouped by their association with *vat*-positive (Vat+) or *vat*-negative (Vat-) strains. The presence of IgG-specific antibodies was determined by ELISA, and compared to results obtained from 42 healthy volunteers with no recent history of UTI (Healthy). A significantly higher IgG titre was observed in plasma of patients infected with Vat+ strains compared to patients infected with Vat- strains and healthy controls.

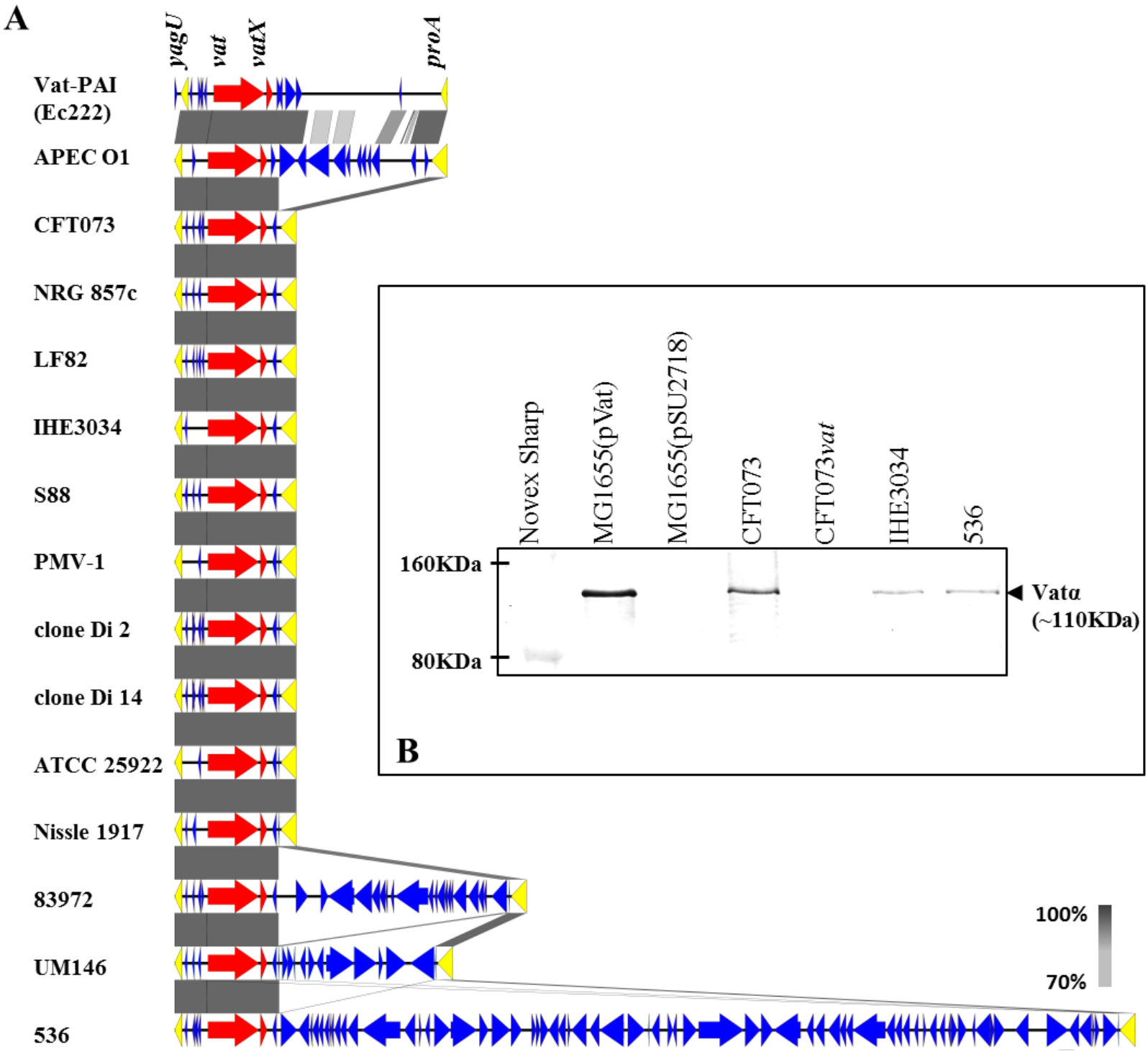
Supplementary Information

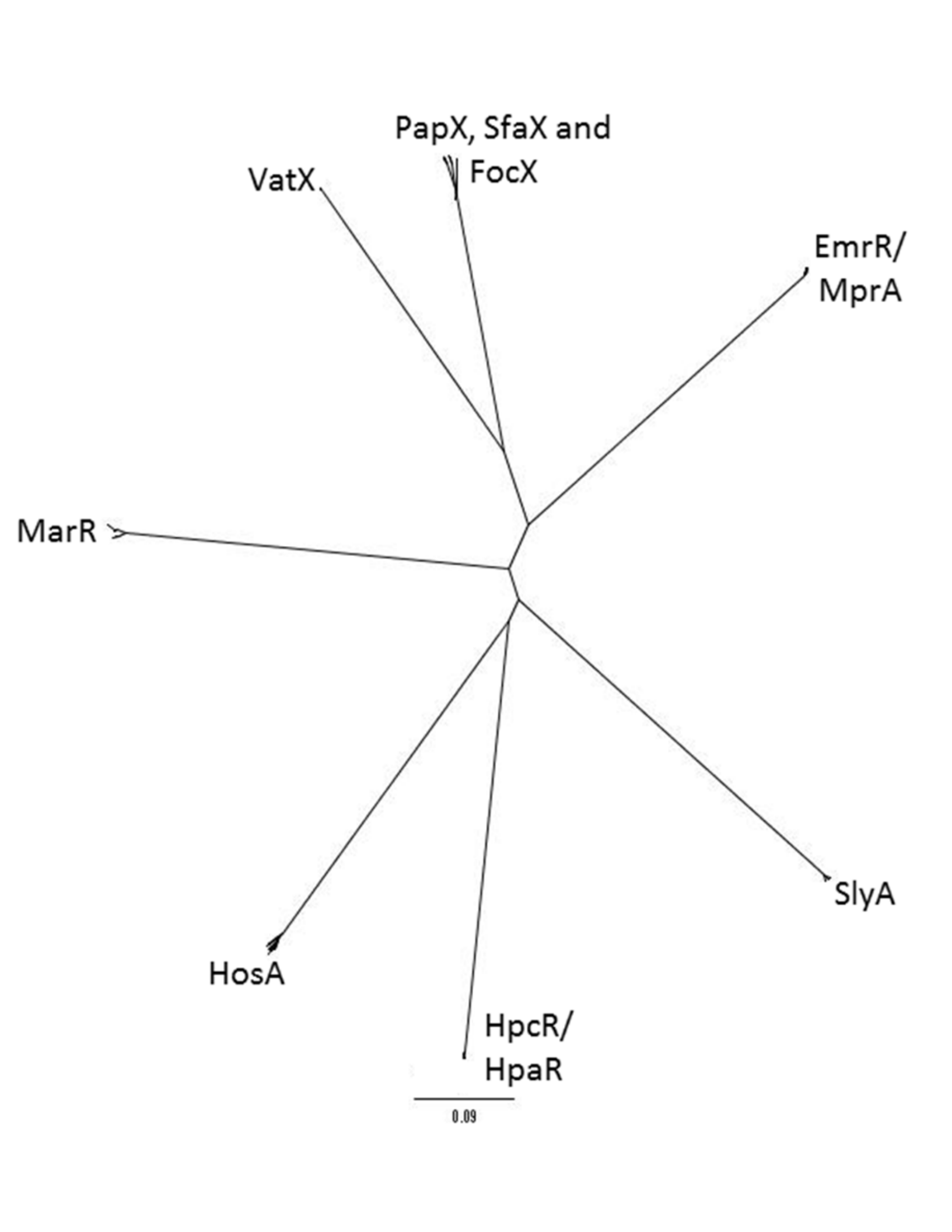
Figure S1. Cladogram demonstrating the relationship of the 330 MarR-type regulator protein sequences identified in the 77 complete *E. coli* genomes listed in Table S1. The scale represents the number of amino acid substitutions per site over 194 positions.

Figure S2. Vat catalytic triad and VIR1/2 regions mapped using the crystal structure of hemoglobin protease (Hbp) passenger domain (3AK5). Hbp is the most related SPATE to Vat, sharing 79% amino acid identity. The structural protein mapping indicates that VIR1/2 are not directly associated with the globular catalytic triad.

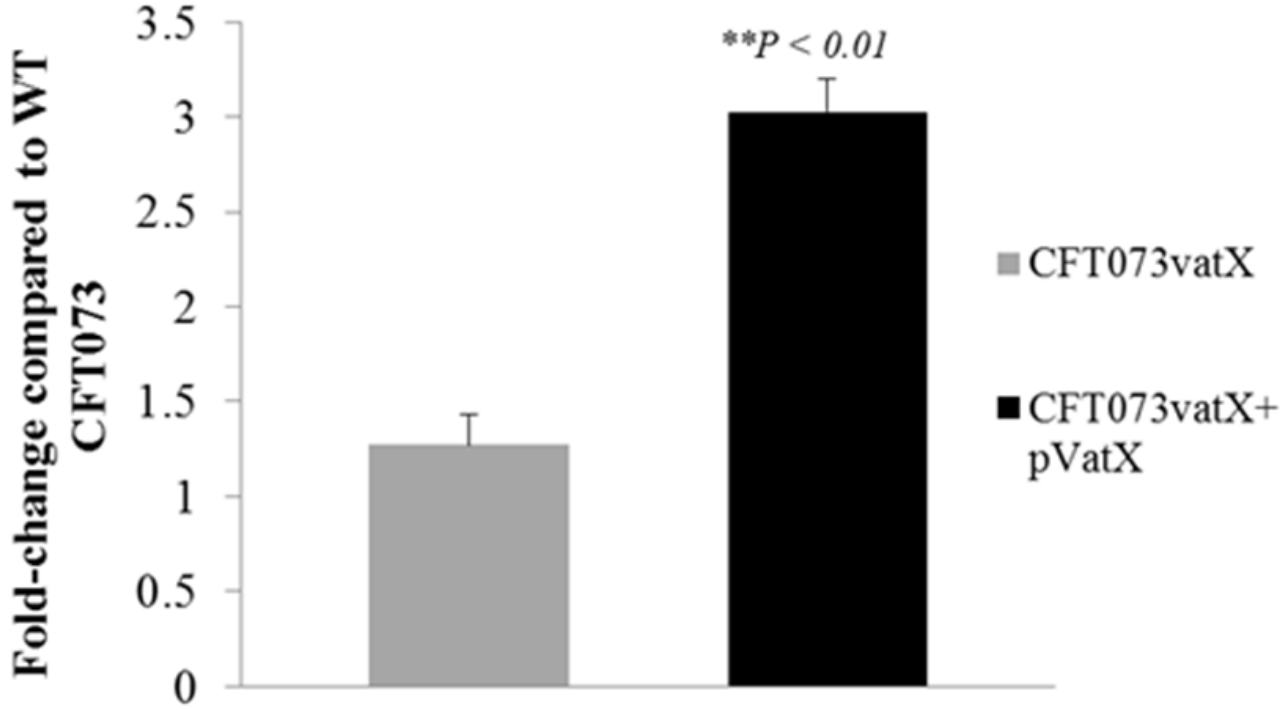
Table S1. List of the 77 sequenced *E. coli* genomes obtained from the NCBI website. The strains include a selection of environmental, non-pathogenic (NP) and pathogenic *E. coli*. The list includes the following *E. coli* pathotypes: enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), adherent-invasive *E. coli* (AIEC), enterohaemorrhagic *E. coli* (EHEC), enteroaggregative haemorrhagic *E. coli* (EAHEC), Shiga toxin-producing *E. coli* (STEC), neonatal meningitis *E. coli* (NMEC), uropathogenic *E. coli* (UPEC) and avian pathogenic *E. coli* (APEC).

Table S2. MarR-type transcriptional regulator genes identified in the 77 complete *E. coli* genomes described in Table S1. The representative genes used as query sequences in the BLAST analysis are underlined. These sequences were used to generate the phylogram in Figure 2. Seven major clades were identified, MarR; MprA/EmrR; HosA; HpcR/HpaR; SlyA; SfaX/FocX/PapX and VatX. The level of amino acid sequence identity for proteins in each clade is indicated.

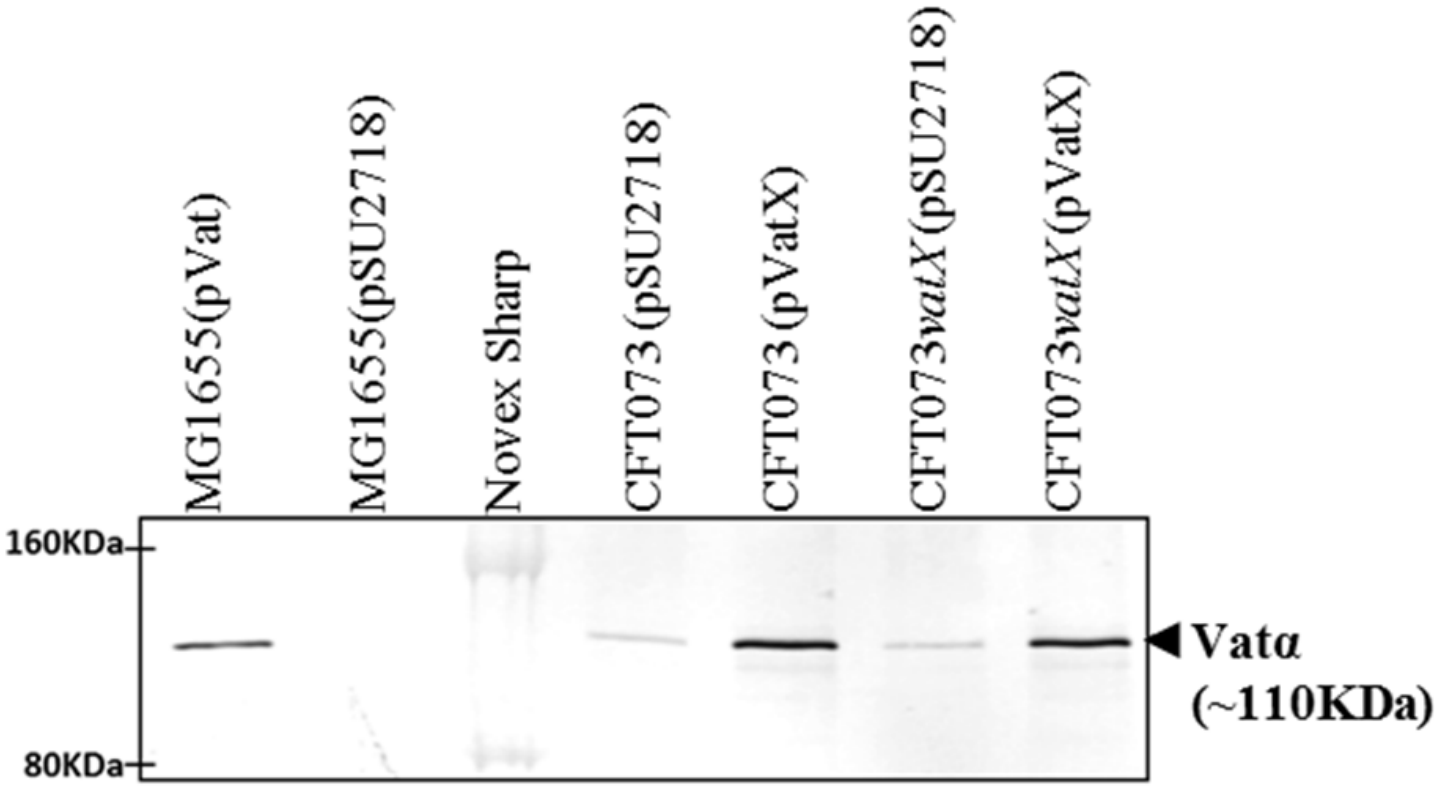

A
yagU
vat
vatX
proA
Vat-PAI (Ec222)
APEC O1
CFT073
NRG 857c
LF82
IHE3034
S88
PMV-1
clone Di 2
clone Di 14
ATCC 25922
Nissle 1917
83972
UM146
536
100%
70%
B
Novex Sharp
MG1655(pVat)
MG1655(pSU2718)
CFT073
CFT073vat
IHE3034
536
160KDa
80KDa
Vatα
(~110KDa)

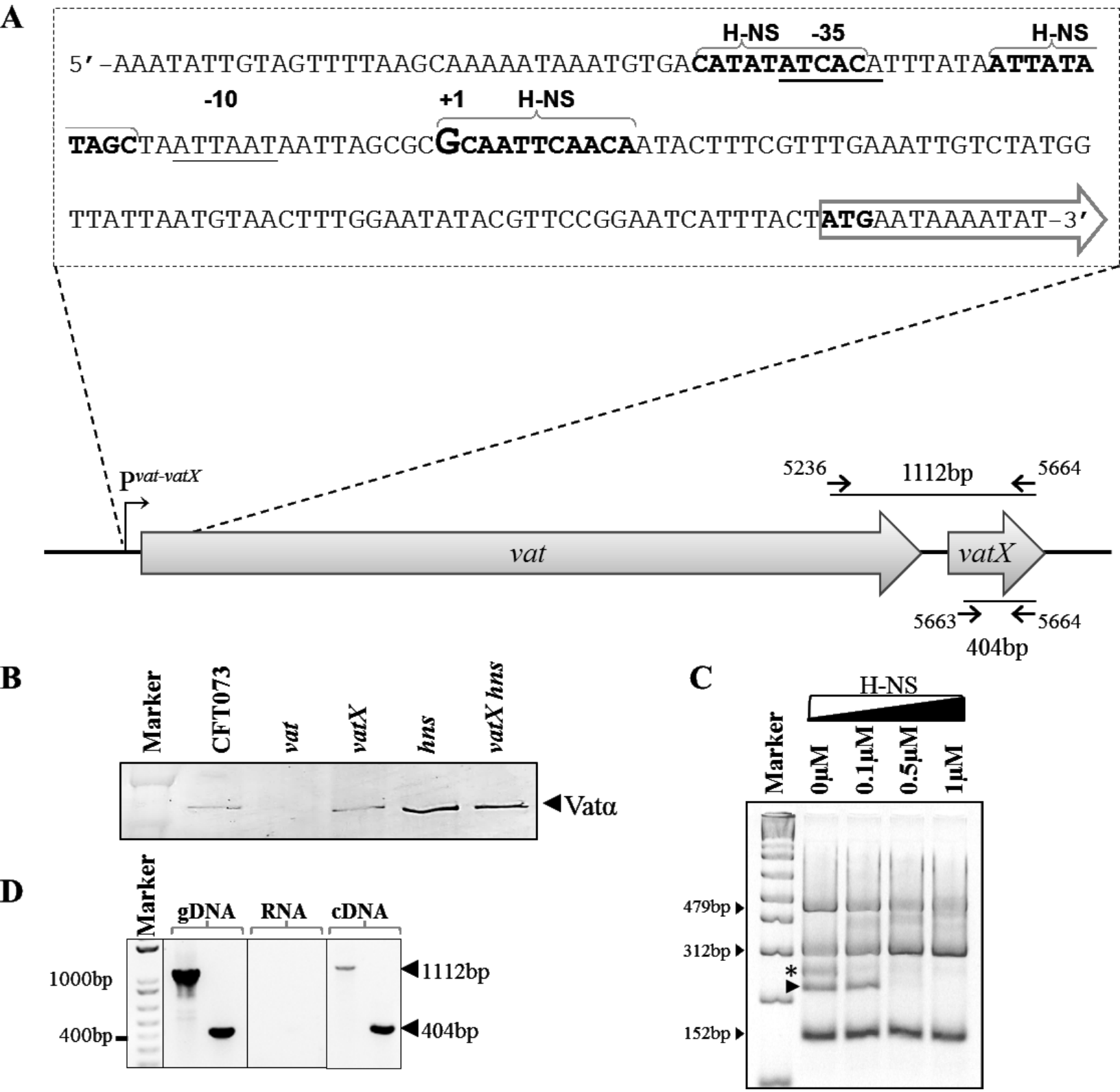

PapX, SfaX and
FocX
VatX
EmrR/
MprA
MarR
SlyA
HosA
HpcR/
HpaR
0.09

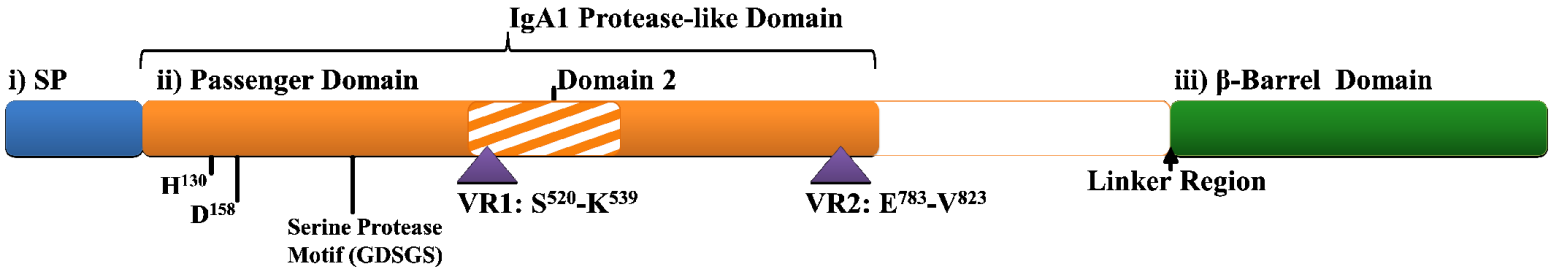
A


Fold-change compared to WT CFT073
3.5
3
2.5
2
1.5
1
0.5
0
**P < 0.01
CFT073vatX
CFT073vatX+ pVatX

B


MG1655(pVat)
MG1655(pSU2718)
Novex Sharp
CFT073 (pSU2718)
CFT073 (pVatX)
CFT073vatX (pSU2718)
CFT073vatX (pVatX)
160KDa
80KDa
Vatα
(~110KDa)

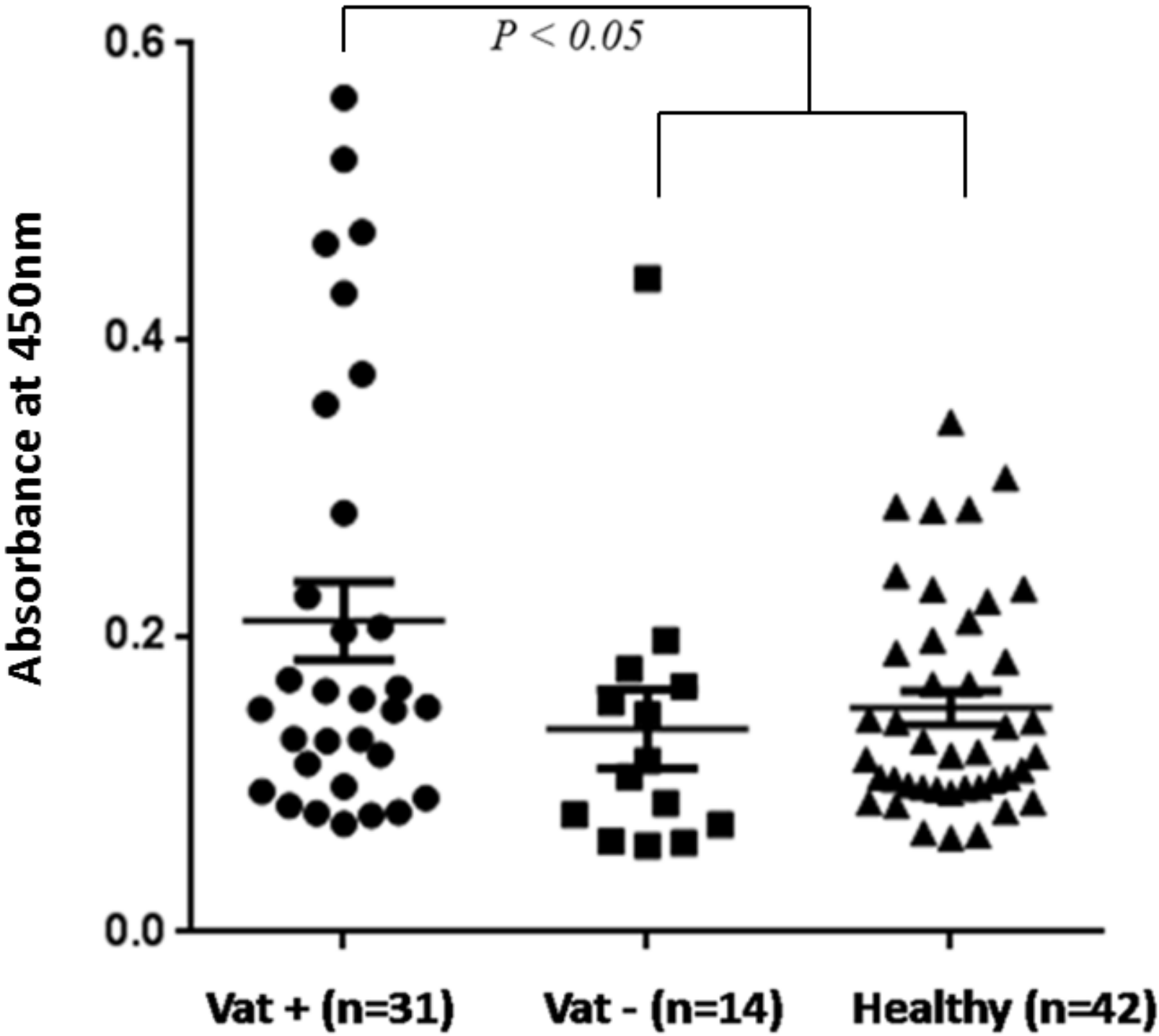

A
H-NS
-35
H-NS
5′-AAATATTGTAGTTTTAAGCAAAAATAAATGTGACATATATCACATTTATAATTATA
-10
+1
H-NS
TAGCTAATTAATAATTAGCGCGCAATTCAACAATACTTTCGTTTGAAATTGTCTATGG
TTATTAATGTAACTTTGGAATATACGTTCCGGAATCATTTACTATGAATAAAATAT-3′
P vat-vatX
5236
1112bp
5664
vat
vatX
5663
5664
404bp
B
Marker
CFT073
vat
vatX
hns
vatX hns
Vatα
C
H-NS
Marker
0μM
0.1μM
0.5μM
1μM
479bp
312bp
*
152bp
D
Marker
gDNA
RNA
cDNA
1000bp
400bp
1112bp
404bp



A

Strain	VR1	VR2	ST	Vat
CFT073	SSDKTANILTLDYQTRPADVK	EIFNGGIQANNSTVNISSDSAVLENSTLTSTALNLNKGANV	ST 73	+
PA48B	.....	.....	ST 73	+
PA10B	.....N..	.....	ST 95	+
PA38B	.....	D.....I.G.....A	ST 537	+
PA32B	.....H.....	D....L.....I.G.....A	ST Unknown	+
PA60B	.....H.....	D.....I.G.....A	ST Unknown	+
PA15B	.....H.....	D.....I.G.....A	ST 12	+
PA56B	.....H.....	D.....I.G.....A	ST 2800	+
PA42B	.....H.....	.....A	ST 420	+

B


P < 0.05
0.6
0.4
0.2
0.0
Absorbance at 450nm
Vat + (n=31)
Vat - (n=14)
Healthy (n=42)